Early Warning Systems: The Neglected Importance of Timing

Joshua A Rolnick, MD, JD, MS[1,2,3,4]*; Gary E Weissman, MD, MS[4,5]

[1]Division of General Internal Medicine, Department of Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania; [2]National Clinician Scholars Program, University of Pennsylvania, Philadelphia, Pennsylvania; [3]Corporal Michael J. Crescenz VA Medical Center, Philadelphia, Pennsylvania; [4]Palliative and Advanced Illness Research Center, Department of Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania; [5]Division of Pulmonary, Allergy, and Critical Care Medicine, Department of Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania.

Automated early warning systems (EWSs) use data inputs to recognize clinical states requiring time-sensitive intervention and then generate notifications through different modalities to clinicians. EWSs serve as common tools for improving the recognition and treatment of important clinical states such as sepsis. However, despite the early enthusiasm, these warning systems have often yielded disappointing outcomes. In sepsis, for example, EWSs have shown mixed results in clinical trials, and concerns regarding the overuse of EWSs in diagnosing sepsis have grown.[1-4] We argue that inattention to the importance of timing in EWS training and evaluation provides one reason that EWSs have underperformed. Thus, to improve care, a warning system must not only identify the clinical state accurately, but it must also do so in a sufficiently timely manner to implement the associated interventions, such as administration of antibiotics for sepsis. Although the literature has occasionally highlighted the importance of timing in electronic surveillance systems, no one has linked the temporal dependence of performance metrics and intervention feasibility to the failure of such warning systems and explained how to operationalize timing in their development.[5-8] Using sepsis as an example, we explain why timing is important and propose new metrics and strategies for training and evaluating EWS models. EWSs are divided into two types: detection systems that recognize critical illnesses at a particular moment and prediction systems that estimate risk of deterioration over varying time frames.[9] We focus primarily on detection systems, but our analysis is also important for prediction systems, which we will discuss in the last section.

CLINICAL TIME ZERO AND POSITIVE PREDICTIVE VALUE

EWS metrics have evolved from focusing on crude measures of discrimination to more clinically relevant metrics, such as the positive predictive value (PPV). The common performance metrics, including the c-statistic, evaluate the performance of EWSs in distinguishing events from nonevents, such as the presence or absence of sepsis in hospitalized patients. How-

ever, the c-statistic does not account for disease prevalence. A given c-statistic is compatible with a wide range of PPVs; a low PPV may limit an EWS's usefulness to promote interventions and generate increased alert fatigue.[10]

However, the PPV, although important, provides no information on the timing of state recognition in relation to clinical time zero. Time zero is the first moment at which a critical state can be recognized based on available data and current medical science. Different approaches, including laboratory values, clinical assessments, retrospective chart reviews, triage times, and others, have been used to measure time zero.[8,11-13] All these approaches feature advantages and disadvantages; the evaluation of timing will exhibit sensitivity to the approach used.[14] Further work is needed to gain additional insights into the measurement of time zero.

Just as the same c-statistic is consistent with varying PPVs, so too is the same PPV consistent with different timing in relation to clinical time zero (Figure). An alert-level PPV of 50% indicates that 50% of the alerts signify true cases of sepsis. However, such a value could also indicate any of the following:

a) 50% true cases of sepsis, with a mean time of 35 minutes after clinical time zero;

b) 50% true cases, with a mean time of 60 minutes before clinical time zero (prediction EWS);

c) 50% true cases of sepsis, with a mean time of 1.3 days since clinical time zero, but with 70% of these cases undiagnosed at the time of EWS detection;

d) 50% true cases of cases, with mean time of 1.3 days since clinical time zero, that is, all cases among those promptly detected and treated through routine clinician oversight.

Each of these situations features differing clinical utility to help meet the hospital objective of increasing early administration of antibiotics. More generally, three dimensions of timing are important for detection systems. The first dimension is the timing of detection relative to time zero. The second is the timing relative to "real-world" clinician detection. The third is timing with respect to the associated clinical objective. For a given PPV, an EWS performs better when detecting a state (1) at, near, or in advance of time zero, (2) prior to clinician detection, and (3) sufficiently in advance of an operational objective to promote change. On the other hand, when an EWS consistently sends alerts after clinician action, it serves a lesser purpose and risks causing alert fatigue; such cases have been described in studies.[15]

© 2019 Society of Hospital Medicine DOI 10.12788/jhm.3229
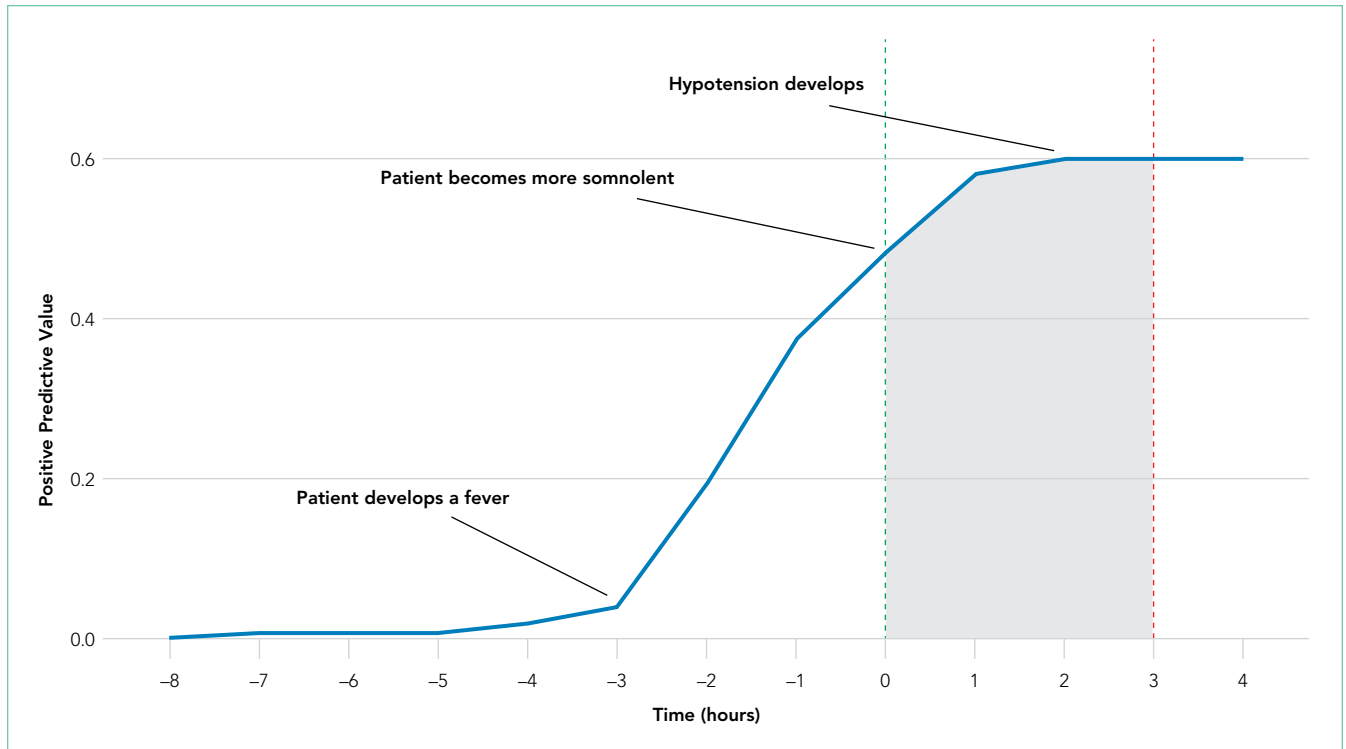
FIG. The Positive Predictive Value Relative to the Evolution of Sepsis. The PPV changes with sepsis evolution, as more information becomes available. The green dotted line depicts clinical time zero, that is, the first point at which sepsis could be recognized based on available data and current medical science. The red dotted line depicts the end of a "clinically important window" based on the operational objective associated with the early warning system.

OPERATIONALIZING TIMING IN EWS TRAINING AND EVALUATION

Acknowledging the importance of timing features implications for researchers and health system leaders. Researchers who develop EWS should include how these systems perform relative to both time zero and critical milestones in the clinical course. Operational leadership should understand the trade-offs that occur between alert fatigue (through lower PPV at the margin with earlier detection) and lead time to implement an intervention. Navigating these trade-offs involves a complex organizational decision. The "number needed to evaluate" is one way to quantify this fatigue factor.[16] Such a measure gives a sense of the number of cases a clinician will need to evaluate per event. Collaborations between clinical leadership, operational leadership, and data scientists are needed to determine how to evaluate individual systems.

A good metric should capture the three important dimensions of timing while retaining intuitiveness to clinicians and leadership. One graphical option involves plotting the PPVs over time and relative to the clinical state evolution (Figure). This PPV-over-time curve shows when true positives occur relative to the time course of sepsis, including the three major dimensions of timing. This curve can also show a clinically important window (CIW), which is bounded on the right by the latest point in time when recognition could still meet the clinical objective. For sepsis, the curve might be bounded at 2.5 hours to meet an objective of antibiotics within three hours, with the assumption that 0.5 hour is needed for a response. For detection systems, the

window would be bounded on the left by clinical time zero. The graph can also designate the point when most cases of sepsis have been recognized clinically with historical data. The Figure depicts an example curve for a detection model.

The metrics derived from this curve may be used alongside the PPV for training and evaluation. Often, adjusting the PPV for its relationship to time zero and the CIW will aid in recognizing the existence of a time beyond which detection fails to help achieve the intended intervention. Detection beyond the window should not credited as a true positive if it fails to facilitate the objective. One option is to credit detection at or before time zero as one and discount later detection by the delay from time zero. More specifically, a true positive could be discounted by the difference between the end of the CIW and the moment of detection divided by the CIW length. This discounted PPV could be displayed alongside the PPV to gauge the temporal dimension of performance and be used for training.

The use of timing places additional demands on validation owing to the need for a time-based gold standard. In such a case, the unit of analysis in system development might not be the patient encounter but rather the patient-hour or patient-15-minute epoch, depending on how frequently the EWS updates risk information and may alert. By contrast, the sepsis detection models used in administrative databases rely on an encounter-level PPV, which provides more limited information compared with real-time EWSs.[17] When time zero cannot be measured, alternatives may be used to capture several dimensions of timing; these alternatives include measurement

of the percentage of cases that recognize the event prior to clinicians.[15]

MOVING TOWARD PREDICTION

Detection systems face the limitation that they lack the capability to identify a state before its occurrence. Prediction systems are more likely to be actionable, as they provide more lead time for intervention, but accurate prediction models are also more difficult to develop. With a predictive system, an additional dimension of timing becomes important: the time horizon for prediction. Prediction models may be trained to recognize a state within a specific time frame (eg, 6, 12, or 24 hours), and test characteristics, including PPV, may vary with the window.[18] A given PPV (of eventual development of sepsis) is compatible with varying time windows and thus again lacks important information on performance.

The timing relative to clinical time zero remains important for prediction. For a predictive EWS, the graph in the figure may be expected to shift to the left. Models with good performance will occasionally send an alert after time zero. For a prediction system with a time horizon of six hours, it is more useful to have alerts occur a mean time of four hours prior to time zero than four minutes prior.

CONCLUSION

Improving the clinical utility of EWSs requires better measurement of timing. Researchers should incorporate timing into system development, and operational leaders should be cognizant of timing during implementation. Specific steps should include devising better strategies to estimate the relationship of state recognition to clinical time zero and developing methods to discount recognition when it occurs too late to be actionable.

References

1. The Lancet Respiratory Medicine. Crying wolf: the growing fatigue around sepsis alerts. *Lancet Respir Med.* 2018;6(3):161. doi: 10.1016/S2213-2600(18)30072-9.
2. Hooper MH, Weavind L, Wheeler AP, et al. Randomized trial of automated, electronic monitoring to facilitate early detection of sepsis in the intensive care unit. *Crit Care Med.* 2012;40(7):2096-2101. doi: 10.1097/CCM.0b013e318250a887.
3. Nelson JL, Smith BL, Jared JD, et al. Prospective trial of real-time electronic surveillance to expedite early care of severe sepsis. *Ann Emerg Med.* 2011;57(5):500-504. doi: 10.1016/j.annemergmed.2010.12.008.
4. Umscheid CA, Betesh J, VanZandbergen C, et al. Development, implementation, and impact of an automated early warning and response system for sepsis. *J Hosp Med.* 2015;10(1):26-31. doi: 10.1002/jhm.2259.
5. Kleinman KP, Abrams AM. Assessing surveillance using sensitivity, specificity and timeliness. *Stat Methods Med Res.* 2006;15(5):445-464. doi: 10.1177/0962280206071641.
6. Jiang X, Cooper GF, Neill DB. Generalized AMOC curves for evaluation and improvement of event surveillance. *AMIA Annu Symp Proc.* 2009;281-285.
7. Futoma J, Hariharan S, Sendak M, et al. An improved multi-output Gaussian process RNN with real-time validation for early sepsis detection. In Proceedings of the 2nd Machine Learning for Healthcare Conference (MLHC), Boston, MA, Aug 2017.
8. Rolnick J, Downing N, Shepard J, et al. Validation of test performance and clinical time zero for an electronic health record embedded severe sepsis alert. *Appl Clin Inform.* 2016;7(2):560-572. doi: 10.4338/ACI-2015-11-RA-0159.
9. DeVita MA, Smith GB, Adam SK, et al. "Identifying the hospitalised patient in crisis"—A consensus conference on the afferent limb of rapid response systems. *Resuscitation.* 2010;81(4):375-382. doi: 10.1016/j.resuscitation.2009.12.008.
10. Romero-Brufau S, Huddleston JM, Escobar GJ, et al. Why the C-statistic is not informative to evaluate early warning scores and what metrics to use. *Crit Care.* 2015;19(1):284-290. doi: 10.1186/s13054-015-0999-1.
11. Evans IVR, Phillips GS, Alpern ER, et al. Association between the New York sepsis care mandate and in-hospital mortality for pediatric sepsis. *JAMA.* 2018;320(4):358-367. doi: 10.1001/jama.2018.9071.
12. Daniels R, Nutbeam T, McNamara G, et al. The sepsis six and the severe sepsis resuscitation bundle: a prospective observational cohort study. *Emerg Med J.* 2011;28(6):507-512. doi: 10.1136/emj.2010.095067.
13. Paul R, Melendez E, Wathen B, et al. A quality improvement collaborative for pediatric sepsis: lessons learned. *Pediatr Qual Saf.* 2018;3(1):1-8. doi: 10.1097/pq9.0000000000000051.
14. Rhee C, Brown SR, Jones TM, et al. Variability in determining sepsis time zero and bundle compliance rates for the centers for medicare and medicaid services SEP-1 measure. *Infect Control Hosp Epidemiol.* 2018;39(9):994-996. doi: 10.1017/ice.2018.134.
15. Winter MC, Kubis S, Bonafide CP. Beyond reporting early warning score sensitivity: the temporal relationship and clinical relevance of "true positive" alerts that precede critical deterioration. *J Hosp Med.* 2019;14(3):138-143. doi: 10.12788/jhm.3066.
16. Dummett BA, Adams C, Scruth E, et al. Incorporating an early detection system into routine clinical practice in two community hospitals: Incorporating an EWS into practice. *J Hosp Med.* 2016;11(51):S25-S31. doi: 10.1002/jhm.2661.
17. Jolley RJ, Quan H, Jetté N, et al. Validation and optimisation of an ICD-10-coded case definition for sepsis using administrative health data. *BMJ Open.* 2015;5(12):e009487. doi: 10.1136/bmjopen-2015-009487.
18. Wellner B, Grand J, Canzone E, et al. Predicting unplanned transfers to the intensive care unit: a machine learning approach leveraging diverse clinical elements. *JMIR Med Inform.* 2017;5(4):e45. doi: 10.2196/medinform.8680.